Mathematical Biosciences
and Engineering

Research article

Cross-subject EEG-based emotion recognition through dynamic optimization of random forest with sparrow search algorithm

Xiaodan Zhang[1], Shuyi Wang[1,*], Kemeng Xu[1], Rui Zhao[1] and Yichong She[2]

[1] School of Electronics and Information, Xi'an Polytechnic University, Xi'an, Shaanxi 710060, China
[2] School of Life Sciences, Xi Dian University, Xi'an, Shaanxi 710126, China

* **Correspondence:** Email: 18632569085@163.com.

Abstract：The objective of EEG-based emotion recognition is to classify emotions by decoding signals, with potential applications in the fields of artificial intelligence and bioinformatics. Cross-subject emotion recognition is more difficult than intra-subject emotion recognition. The poor adaptability of classification model parameters is a significant factor of low accuracy in cross-subject emotion recognition. We propose a model of a dynamically optimized Random Forest based on the Sparrow Search Algorithm (SSA-RF). The decision trees number (DTN) and the leave minimum number (LMN) of the RF are dynamically optimized by the SSA. 12 features are used to construct feature combinations for selecting the optimal feature combination. DEAP and SEED datasets are employed for testing the performance of SSA-RF. The experimental results show that the accuracy of binary classification is 76.81% on DEAP, and the accuracy of triple classification is 75.96% on SEED based on SSA-RF, which are both higher than that of traditional RF. This study provides new insights for the development of cross-subject emotion recognition, and has significant theoretical value.

Keywords: cross-subject; emotion recognition; SSA-RF; DTN; LMN

1. Introduction

Research in neuroscience and psychology has shown that EEG signals can intuitively reflect an individual's emotional changes [1]. EEG signals are subject to individual differences and are non-stationarity [2], so the construction of a cross-subject emotion recognition model has become an important research direction and has great significance. This study focuses on exploring the emotions

generated by the same emotion-inducing mode between subjects and cross-subjects, and conducts training based on cross-subjects' emotional characteristics, aiming to improve the emotion classification accuracy of the cross-subjects emotion recognition model [3].

Machine learning is the method of making computers with human intelligence, which requires training models to improve themselves by learning from data, a technique that is well suited to tasks such as processing electrical brain signals. Aljuhani et al. [4] used machine learning algorithms to identify emotions from speech, extracting various spectral features, such as the Mel-frequency cepstrum coefficient (MFCC) and mel spectrum, and obtained 77.14% accuracy by the SVM method. Liu and Fu [5] trained a support vector machine in emotion recognition and proposed a multi-channel feature fusion method. The recognition accuracy of different subjects ranged from 0.70 to 0.87, and the results of PLCC and SROSS measurements reached 0.843 and 0.789. Salido Ortega et al. [6] used machine learning technology to establish individual models, general models, and gender models to automatically identify subjects' emotions, which verified that their individual emotions are highly correlated with the situation. They used the situation data to realize automatic recognition of emotions in real situations. Karbauskaite et al. [7] studied facial emotion recognition, and combining four features made the emotion classification accuracy reach 76%. Xie et al. [8] proposed a transformer-based cross-mode fusion technology and blackmail network architecture for emotion estimation, and this multi-mode network architecture can achieve an accuracy of 65%. Li et al. [9] proposed a TANN neural network. Adaptive highlighting of transferable brain region data and samples through local and global attention mechanisms was used to learn emotion discrimination information.

Deep learning is a technique that combines low-level features to form more abstract high-level features or categories, so as to learn effective feature representations from a large number of input data and apply these features to classification, regression, and information retrieval. It is also applicable to the processing of EEG signals. Jiang et al. [10] established a 5-layer CNN model to classify EEG signals, and the average accuracy reached 69.84%, 0.79% higher than that of the CVS system. Zhang and Li [11] proposed a teaching speech emotion recognition method based on multi-feature fusion deep learning, and the recognition accuracy reached 75.36%. Liu and Liu [12] applied BP (back-propagation) neural network as Technical Support and Combines EEG Signals to Classify Criminal Psychological Emotions. Liu et al. [13] used the MHED dataset to study the multi-modal fusion network of video emotion recognition based on hierarchical attention, and the accuracy was 63.08%. Quan et al. [14] showed that interpersonal characteristics can help improve the performance of automatic emotion recognition tasks, and the highest accuracy at the titer level was 76.68%. Fang et al. [15] proposed a Multi-feature Deep Forest (MFDF) model to identify human emotions.

We employed the random forest (RF) classification model in the field of machine learning. As the integration of decision trees, RF is a classifier that uses multiple trees to train and predict samples. It has the advantages of being built easily, able to obtain the importance weight of features, and is less likely to overfit. Anzai et al. [16] used the machine learning random forest algorithm to build a fragile classifier and a descent classifier to identify the frail state and fall risk of the elderly, and the overall balance accuracy for the identification of frail subjects was $75\% \pm 0.04\%$. The overall balance accuracy for classifying subjects with a recent history of falls was 0.57 ± 0.05 (F1 score: 0.62 ± 0.04).

In the field of optimization classification model, related researchers have made great progress. Zhang et al. [17] used Bayesian super parameters to optimize the stochastic forest classifier on Sentinel-2 satellite image urban land cover classification. As a result, the RF after Bayesian optimization was 0.5% higher than RF by using RGB band features, and its accuracy increased 1.8%

by using multi-spectral band features. Beni and Wang [18] proposed swarm intelligence in 1989. The probabilistic search algorithm built by simulating the swarm behavior of natural organisms was intelligent, because it was independent of the optimization problem itself, required fewer parameters, had high fault tolerance, and had strong stability. Ye et al. [19] adopted a genetic algorithm to optimize the decision tree combination in the parametric optimization random forest, comparing with the actual profit, the profit score of RFoGAPS increases by 7.73%.

In recent years, the swarm intelligence algorithm based on biological characteristics has been widely used in electronic information, engineering technology, biomedicine, and other fields. Sparse Bayesian Learning for end-to-end spatio-temporal-filtering-based single-trial EEG classification (SBLEST) optimized spatio-temporal filters and the classifier simultaneously within a principled sparse Bayesian learning framework to maximize prediction accuracy [20,21]. Since feature extraction and emotion classification were completed independently at different stages in the EEG decoding process, and the research aimed to reduce the cost generated in the classification process, we put forward an optimization method that can dynamically optimize the parameters of RF model, which can improve the accuracy. At the same time, the intelligent optimization algorithm we sought should be as simple in structure as possible, easy to implement, and with few control parameters, so we selected the Sparrow Search Algorithm (SSA). We applied the SSA to optimize the key parameters of RF and improve the classification accuracy in cross-subject emotion recognition.

SSA-RF was used on the DEAP and SEED datasets, which verified that it had better adaptability, effectiveness, necessity of classification model parameters, and reduced subject dependency.

2. Methods

2.1. Preprocessing

Windowing is employed to avoid overfitting which is caused by small data. For data of $T(s)$, the time window is $m(s)$, and the overlap rate is 50%. The principle of windowing is shown in Figure 1.
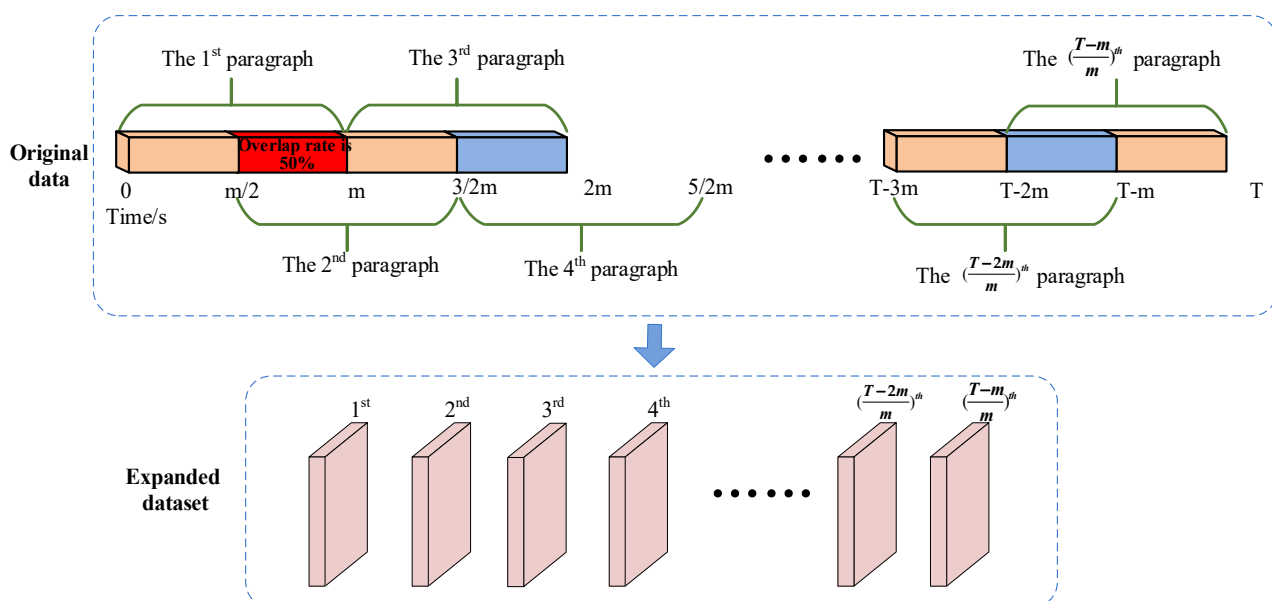

Figure 1. The principle of windowing.

2.2. Feature extraction

Feature extraction is one of the crucial processing components of cross-subject emotion recognition, which can mine the hidden information of mental activity and cognitive function.

Different emotional features are reflected in different physical quantities of signals. The wavelet transform is effective in finding the optimal trade-off between time and frequency resolution. Traditional features are extracted from the time-domain, frequency-domain, and time-frequency domains [22]. Soroush et al. [23] obtained good classification accuracy by applying the characteristics of mean, skewness, and Shannon entropy. The research motivation comes from the combinations of different features [24] or principal component analysis and discrete wavelet transform for feature selection [25].

In this paper, 9 features of time domain, 2 features of frequency domain, and 1 feature of time-frequency domain are extracted for SSA-RF cross-subject emotion recognition. We used all channels, which can provide more information.

2.2.1. Time domain features

In the time domain, the zero crossing rate (ZCR), standard deviation (SD), mean, root mean square (RMS), energy (Eng), skewness, approximate entropy (ApEn), sample entropy (SampEn), and Hjorth are extracted as the features of the EEG, which are shown in Table 1.

2.2.2. Frequency domain features

We transform the time domain EEG to the frequency domain through the Fast Fourier Transform (FFT), and the Power Spectral Density (PSD) and Differential Entropy (DE) are extracted as the features, which are shown in Table 2.

2.2.3. Time frequency domain feature

Due to the characteristics of both time and frequency domains, the time-frequency domain can comprehensively reflect the information of the EEG, which is a more comprehensive display of EEG feature information. After transforming the EEG into the time-frequency domain through the wavelet transform (WT), the wavelet shannon entropy (SE) is extracted as the feature, which provides uncertainty, information content, spectral characteristics, and time-frequency variation characteristics of the EEG, and reveals the correlation between the EEG signal and emotion so as to realize emotion recognition. The SE of the time-frequency domain can describe the information content and complexity of the signal at different times and frequencies, as shown in Eq (1):

$$H(X) = -\sum_x P(x) \, log_2[P(x)] \tag{1}$$

Here, *H(X)* is the SE, and $P(x)$ is the probability value of different sample data, in bits.

2.3. The principle of SSA-RF

SSA is used to obtain the optimal NDT and MNL of the RF dynamically, and the SSA is inspired by the foraging behavior of sparrows to obtain the optimal parameters.

According to the biological rules of the SSA, the discoverer first finds the optimal foraging area. Followers search for food in the area around the discoverers or obtain food from them. They may also engage in food plundering between individuals and update their foraging area. When the sparrows are aware of danger, they also update their foraging area to avoid being attacked by predators. Assuming there are n sparrows in d-dimensional space, X represents the position of the sparrow. The main responsibility of discoverers is to find food for the population and guide their followers in the foraging direction. According to this rule, the location of the discoverer is updated as described in Eq (2):

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot exp\left(\frac{-i}{\alpha \cdot i_{ter\,max}}\right) & R_2 < S_T \\ X_{i,j}^t + Q \cdot L & R_2 \geq S_T \end{cases} \tag{2}$$

Here, $i_{ter\,max}$ is the maximum number of iterations, t is the current number of iterations, X_{ij} is the position information of the sparrow i in the j dimension, R_2 and S_T are the warning and safety values respectively, Q and α ($\alpha \in (0, 1]$) are the random numbers, Q follows a normal distribution, and L is a matrix with all elements being $1 \times d$. When $R_2 < S_T$, there are no predators in the foraging environment, and the discoverer can conduct a safe and extensive search. When $R_2 \geq S_T$, some sparrows confirm the presence of predators and issue an alert, and all the sparrows need to move to the feeding area in a timely manner.

The position update of followers is described in Eq (3):

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot exp\left(\frac{x_{worst}^t - X_{i,j}^t}{i^2}\right) & i > n/2 \\ X_P^{t+1} + \left|X_{i,j}^t - X_P^{t+1}\right| \cdot A^+ \cdot L & other \end{cases} \tag{3}$$

Here, X_P is the optimal position occupied by the current discoverer, and X_{worst} is the current global worst position, A is a matrix of $1 \times d$, with elements randomly assigned to 1 or -1, and $A^+ = A^T(AA^T)^{-1}$. When $i > n/2$，The follower i is in a state of hunger, whose fitness is reduced, and in order to find food, it needs to change areas for foraging.

When aware of danger, the sparrow population will engage in anti-predatory behavior, as described in Eq (4):

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot \left|X_{i,j}^t - X_{best}^t\right| & f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{\left|X_{i,j}^t - X_{worst}^t\right|}{(f_i - f_w) + \varepsilon}\right) & f_i = f_g \end{cases} \tag{4}$$

Here f_i is the fitness value of the current sparrow individual, f_g and f_w are the current global best and worst fitness, X_{best} is the current global optimal position, β represents the wavelength control parameter, which is a random number subject to a standard normal distribution (mean 0, variance 1), K ($K \in [-1,1]$) is a random number representing the direction of sparrow movement, and ε is the smallest constant to avoid a denominator of 0.

When $f_i > f_g$, sparrows are in a hazardous area and can be easily spotted or attacked by natural predators.

When $f_i = f_g$, sparrows realize that they are currently in a dangerous position, and in order to avoid being attacked by predators, they need to move closer to the sparrows in the safe area to reduce the likelihood of predation. The implementation of SSA-RF is shown in Table 3.

Table 1. The 9 features of the time domain.

Feature	Definition	Formula and description	The connection to emotions
ZCR	The number of times the signal passes the zero value in unit time.	$Z_x = \frac{1}{N} Z_{num}(x)$, where N is the length of the signal sample. Z_{num} is the number of times the signal passes through zero in unit time.	ZCR is closely related to positive emotions. The higher the ZCR value, the more significant the positive emotions will be.
SD	The degree of dispersion among individuals in the sample, the amplitude of indirect reaction signal change from the mean.	$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$, where n is the number of samples, x_i is the value of each data, and $\bar{x}$ is the mean of the sample.	SD indirectly represents signal changes from the mean, so as to judge whether the brain activity pulls away from the stationary state.
Mean	Indirectly reflects the intensity of brain activity.	$\mu_\xi = \frac{1}{T}\sum_{t=1}^{T}\xi(t)$, where $\xi(t)$ is the time domain data, and T is the data length.	Mean represents the intensity of brain activity in a certain period of time, the higher the value, the greater the intensity of activity.
RMS	The degree to which the data value of each frame of the EEG signal deviates from the mean value of the overall sample signal.	$rms = \sqrt{\frac{\sum_{i=0}^{n}x_i^2}{n}}$, where x_i is the time domain signal data, and n is the sample length.	RMS represents the degree to which the data value of each frame deviates from the mean of the whole sample signal, reflecting the degree of deviation from the intensity of brain activity.
Eng	EEG is variable and non-stationary, and its total energy is infinite.	$E_x = \int_{-\infty}^{\infty}|x(t)|^2 \, dt$, where $x(t)$ represents the signal data value at a certain time, and the total energy is the integral of the square of the signal data.	Eng can capture the emotional change and evolution trend, the higher the Eng value, the stronger the positive emotion will be.
Skewness	The distribution symmetry of the values of a particular population is described.	$b_1 = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{3/2}}$, where $\overline{x}$ is the mean, s is the standard deviation, and m_3 is the third-order central matrix.	Skewness represents the degree of deviation between each frame and the normal distribution. The larger the value, the larger the skewness of its distribution form.
ApEn	A nonlinear parameter used to quantify the regularity and unpredictability of time series fluctuation	(1) $X_i = [x(i), x(i+1), \ldots, x(i+x-1)]$ (2) $d[X_i, X_j] = max|x(i+k) - x(j+k)|, k \in (0, m-1)$ (3) $B_i^m(r) = \frac{B_i}{N-m+1}$ (4) $B^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln B_i^m(r)$	ApEn represents the complexity of EEG and reflects the possibility of new information. The more complex time series, the greater the value will be.

Continued on next page

Feature	Definition	Formula and description	The connection to emotions
ApEn	A nonlinear parameter used to quantify the regularity and unpredictability of time series fluctuation	(5) $ApEn(m,\ r,\ N) = B_i^m(r) - B^{m+1}(r)$ (6) Arrange the elements of the time series X in order as vectors with m dimension (7) Define $d[X_i,\ X_j]$ as the distance between vector X_i and vector X_j (8) Write B_i as the number of $d[X_i,\ X_j] \leq r$ (r is the similarity tolerance), and calculate the ratio of B_i to the total number of vectors (N-m+1) (9) Take the logarithmic operation on $B_i^m(r)$, and then find its average of all i, and write it as $B^m(r)$ (10) Make $m = m + 1$ and repeat (1)–(4) to obtain $B^{m+1}(r)$. (11) Final representation approximate entropy	ApEn represents the complexity of EEG and reflects the possibility of new information. The more complex time series, the greater the value will be.
SampEn	The probability of generating new patterns in the sequence when measuring the complexity and dimensional changes of EEG.	(1) $B_i^m(r) = \frac{1}{N-m} num\{d[X_i,\ X_j] < r\}$ (2) $B^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} B_i^m(r)$ (3) It is the same as the approximate entropy in the first two steps. Starting from the third step, the specific steps are as follows: (4) Given threshold $r(r > 0)$, count the number of $d[X_i,\ X_j] < r$ and its ratio to the total number of vectors $(N - m)$ (5) Average the results of the previous step (6) Add dimension m to 1 and repeat the above four steps (7) The actual number of samples is limited, and the final sample entropy is obtained	SampEn measures the probability of generating new patterns in sequence when the EEG complexity and dimension change. The higher the probability, the greater the complexity.
Hjorth	Describe the three time-domain feature sets of EEG single channel, including activity, mobility, and complexity.	$HA = \sigma_0^2;\ HM = \frac{\sigma_1}{\sigma_0};\ HC = \frac{\sigma_2 \sigma_0}{\sigma_1^2}$ where σ_0 is the standard deviation of the signal, and σ_1 and σ_2 are the standard deviations of the first and second derivatives of the signal.	Hjorth represents the EEG changes at different time and spatial locations, thereby revealing the rules and characteristics of brain electrical activity.

Table 2. The 2 features of frequency domain.

Feature	Definition	Formula and description	The connection to emotions
PSD	Represents the conversion of signal strength to unit bandwidth frequency, i.e. the distribution of signal strength within the frequency range.	(1) EEG signals $s[0], s[1], \cdots\cdots, s[N-1]$ are divided into k segments, calculate the windowed discrete Fourier transform $S_{k(v)}$, which is shown as follows: $S_{k(v)} = \sum_m S[m]w[m]exp(-j2\pi vm)$ where, m is between $(k-1)L$ and $M+(k-1)L-1$, $w(m)$ is the window function, M is the segment size, and L is the number of information points between segments, $v = i/M$, where $-\left(\frac{M}{2}-1\right) < i < \frac{M}{2}$. (2) Calculate the modified periodic chart value using the formula: $P_k(v) = 1/wabs\left(S_k(v)\right)^2$ where, $w = \sum_{m=0}^{M} w^2[m]$. (3) Estimate the power spectral density by using the average of the periodic plot values calculated using the equation: $L_s(v) = 1/K \sum_{k=1}^{K} P_k(v)$ where, the number of points shared by two adjacent signal segments is equal to *(M-L)*, which means that the two adjacent segments will be overlapped by *(M-L)* points.	PSD represents the energy distribution of EEG signals in different frequency bands, and identifies emotional states through the difference of energy distribution.
DE	It is a generalization of Shannon's information entropy $-\sum_x p(x) \log\left(p(x)\right)$ on continuous variables.	$$DE = -\int_a^b p(x) \log\left(p(x)\right) dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}}\right) dx$$ $$= \frac{1}{2} \log\left(2\pi e\sigma_i^2\right)$$ Here $p(x)$ represents the probability density function of continuous information, and [a, b] represents the interval of information values, which is equal to the logarithm of its energy spectrum in a specific frequency band.	DE represents the complexity and irregularity of EEG signals in the frequency domain and captures the dynamic changes.

Table 3. The implementation of SSA-RF.

Algorithm: SSA-RF
Input: D: EEG Data
G: Number of iterations
P: Number of population
F_0: Global optimal fitness value
F: Current fitness value
Output: Optimal DTN and LMN of RF
1:   Initialize the RF model and substitute D into it
2:   Determine the initial location of sparrow population
3:   while i < G do
4:        for m = 1 to P do
5:           Use the fitness function to determine the global fitness value
6:           Update population position based on fitness ranking order
7:            if F < F_0 then
8:              Update the global optimal position
9:            end if
10:       end for
11:   Select the global optimal position
12:   end while
13:   Extract the two dimensional data of the global optimal position (DTN and LMN) and substitute it into the RF model to output the results

In SSA-RF, the fitness function is used to search for the optimal number of DTN and LMN. The classification error of the training and testing sets is used as the fitness. After the model training is completed, the optimal position of the sparrow population is output, corresponding to the optimal number of DTN and LMN in the RF, Finally, the optimization results were incorporated into the RF for experimentation, which summarized the complete process of SSA-RF optimization parameters. The flowchart of SSA-RF algorithm is shown in Figure 2.
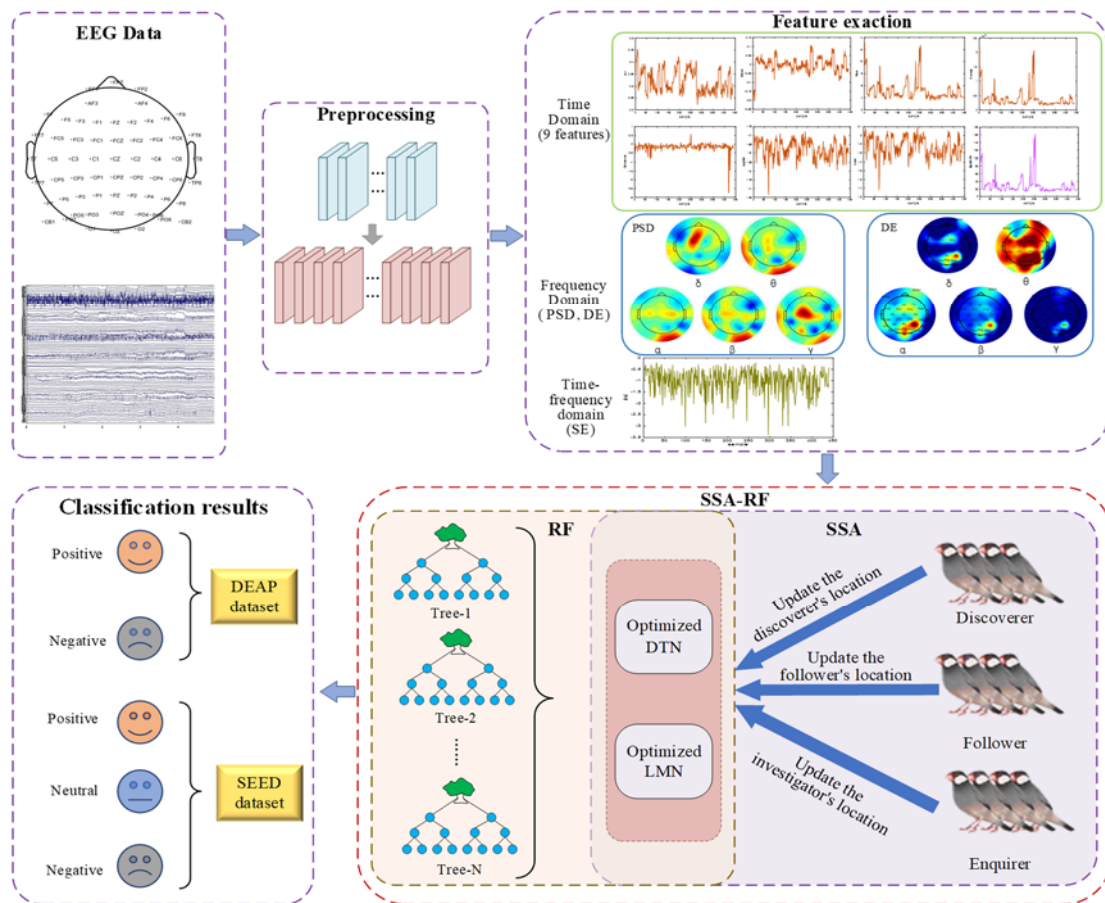
Figure 2. Overview of the SSA-RF flow.

3. Experimental results and analysis

3.1. Datasets and parameters

3.1.1. Datasets

Table 4. Data formats of DEAP and SEED datasets.

Dataset	Data format	Caption
DEAP	40 × 32 × 7680	40: video 32: channel 7680: data
SEED	15 × 3 × 62 × M	15: video 3: number of experiments 62: channel M: data

The DEAP dataset was established by Koelstra et al. [26] from Queen Mary's College London, which included multi-channel physiological signals, facial expression videos, and emotional self-evaluation labels established using the SAM (Self Assessment Manikins) table. It collected EEG

data from 32 healthy subjects (16 males and 16 females), with the first 32 channels of EEG.

The SEED dataset was established by the BCMI laboratory, which recorded the EEG of 15 subjects (including 7 males and 8 females), with an average age of 23.37 years old. Each group had 15 experiments, each consisting of 5 seconds of suggestion before the start, 4 minutes of movie clips, 45 s of self-evaluation, and 15 s of rest. The emotion-inducing materials consist of 15 segments from six movies. After watching the videos, participants recorded their emotional reactions by filling out questionnaires, which were divided into three types: positive emotions, neutral emotions, and negative emotions.

The data formats of the two datasets were shown in Table 4.

3.1.2. Parameters of SSA-RF

The parameters of SSA-RF were population number, maximum number of iterations, dimension, upper boundary, and lower boundary. These parameter values are shown in Table 5.

Table 5. Parameters of SSA-RF.

Parameter	Value
Population number	8
Maximum number of iterations	20
Dimension	2
lower boundary	1
Upper boundary	50

3.2. Preprocessing

For the DEAP dataset, it was composed of participants collecting EEG signals while being emotionally induced, then the participants labeled the label size by watching video through their personal subjective emotions.

The baseline signal mean was removed during baseline processing [27]. The pre-processed data were augmented by windowing. There were 40 sets of data from 40 emotion-inducing videos of each subject, each of which lasted 60 s. These data were processed with a windowing of 10 s and a 50% overlap rate, each video was reconstructed into 440 video data of 10 s. Each video sample had a duration of 10 s, a data sampling rate of 128 Hz, and 32 unaltered channels. Therefore, the amount of data per sample for a single channel was 1280. The original data was reconstructed from $40 \times 32 \times 7680$ to $40 \times 32 \times 14,080$.

For the SEED dataset, it was based on the premise of determining the label to which the emotion-evoking material belongs. In the stage of preprocessing, all data of each subject was integrated and reconstructed, and the original data format of $15 \times 3 \times 62 \times M$ was reconstructed into a format of $225 \times 186 \times M$. The dataset had a total of 15 subjects, and each subject had 15 segments of emotional stimulation materials. All emotional stimulation material data was unified into 15×15, so the first dimension was 225, Each subject was separated for a period of time to perform the same experiment 3 times, each experiment collected 62 channels of EEG signals, and the second dimension was 3×62, which was 186, M was the amount of data in a single channel of each trial,

because each test was not the same time duration, so M was in the range of 37,001–53,001. The original data was reconstructed from $15 \times 3 \times 62 \times M$ to $225 \times 186 \times M$.

3.3. Cross-subject emotion recognition on the DEAP dataset

We conducted 15 randomized grouping experiments. The 25 subjects were randomly selected as the training set, and the other 7 people were selected as the test set for 30 experiments, and 20 iterations were carried out in each experiment.

A total of 20 features were extracted: ZCR, SD, Mean, RMS, Eng, Skew, ApEn, SampEn, Hjorh, PSD, and DE of five frequency bands (δ, θ, α, β, and γ), and SE in the time-frequency domain. We performed many experiments of different feature combinations, and selected the top 8 combinations with high accuracy. There are shown as follows:

Combination 1: All features of the composite domain (20 features);

Combination 2: ZCR, SD, Mean, RMS, Eng, Skewness, ApEn, SampEn, PSD, DE, SE;

Combination 3: SD, RMS, Eng, PSD-δ, DE-δ, DE-β, DE-γ;

Combination 4: F-all, SE, and Hjorth;

Combination 5: Mean, SampEn, DE-β, DE-γ, PSD-β, PSD-γ, SE;

Combination 6: SD, Mean, RMS, Eng, Skewness, Apen, SampEn, DE-α, DE-β, DE-γ, PSD-α, PSD-β, and PSD-γ;

Combination 7: ZCR, SD, Mean, RMS, Eng, Skewness, ApEn, SampEn, DE-α, DE-β, DE-γ, PSD-α, PSD-β and PSD-γ;

Combination 8: SD, Mean, RMS, Eng, and F-α, β, γ;

The value of parameters of RF were generally based on empirical data, and the empirical values of DTN and MLN were 30 and 1, respectively, but they were not suitable for each type of data. We applied SSA algorithm to automatically search for optimal parameters (DTN and MLN) of RF for the 8 combinations, and the optimal values of DTN and MLN are shown in Table 6.

Table 6. Optimal parameter values of different feature combinations based on SSA-RF.

Parameter / Feature combination	1	2	3	4	5	6	7	8
DTN	34	50	49	37	29	24	24	27
MLN	1	1	1	2	2	1	3	1

It can be seen from Table 6 that the optimal parameter values of different feature combinations were different, and they were different from the empirical values, DTN especially showed significant differences. To test which feature combination can achieve the highest accuracy, we experimental with 100 epochs for each feature combination based on the DEAP dataset.

Figure 3 shows the violin plots of the accuracy for different combinations. The median accuracy of combination 3 was higher than the others, and the median accuracy of combination 8 was the lowest. On the whole, the accuracy of each combination was in the range of 72–81%.
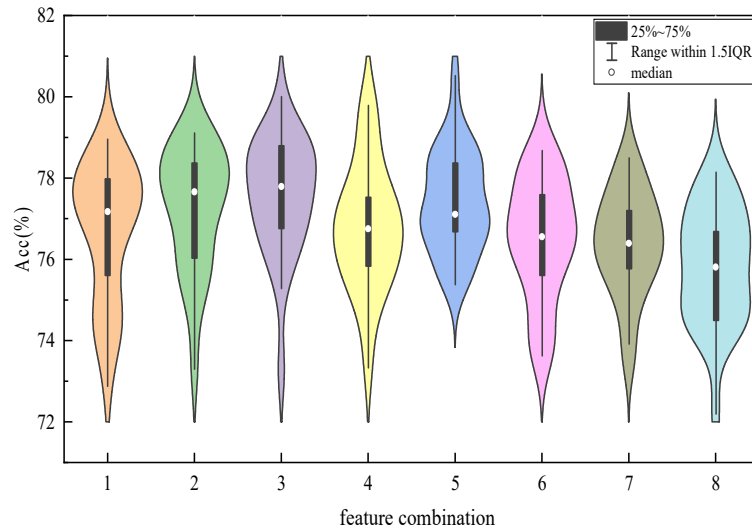
Figure 3. Results of different feature combinations of the DEAP dataset.

The experimental results of SSA-RF on the DEAP dataset showed the accuracy of the test set was improved compared with RF. The classification results and improvement amount are shown in Figure 4 and Table 7.
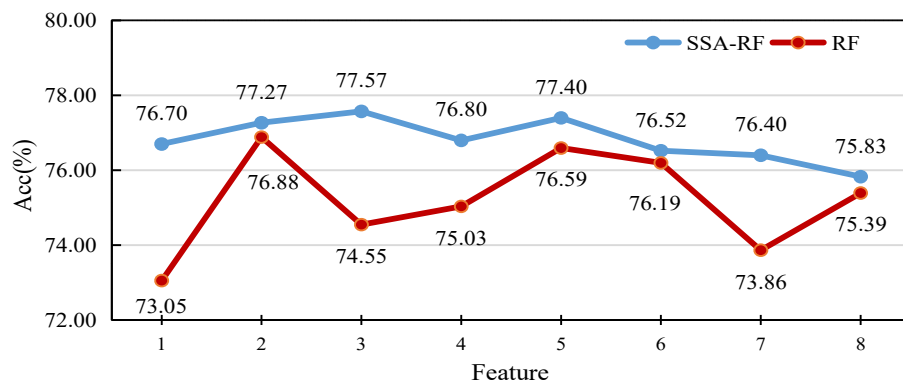

Figure 4. Accuracy of test set based on SSA-RF and RF.

Table 7. The test set recognition accuracy comparison of SSA-RF and RF.

Feature combination Method (%)	1	2	3	4	5	6	7	8
SSA-RF	76.70	77.27	**77.57**	76.80	77.40	76.52	76.40	75.83
RF	73.05	76.88	**74.55**	75.03	76.59	76.19	73.86	75.39
Difference	3.65↑	0.39↑	**3.02↑**	1.77↑	0.81↑	0.33↑	2.54↑	0.44↑
Average				1.62↑				

From Table 7 and Figure 4, it could be seen that the accuracy of SSA-RF was higher than RF on each feature combination, with an average improvement of 1.62%. Among them, combination 1 had

the highest improvement, which was 3.65%, while combination 3 had the highest accuracy with a growth of 3.02%. Combination 3 was selected as the optimal feature combination.

Then, we analyzed the misjudged subjects based on combination 3. Subject 15 was used for analysis that misjudged negative emotions as positive emotions. We compared the features of combination 3 (SD, RMS, Eng, PSD-δ, DE-δ, DE-β, DE-γ) with the mean of the same features in the training set, as shown in Figure 5.
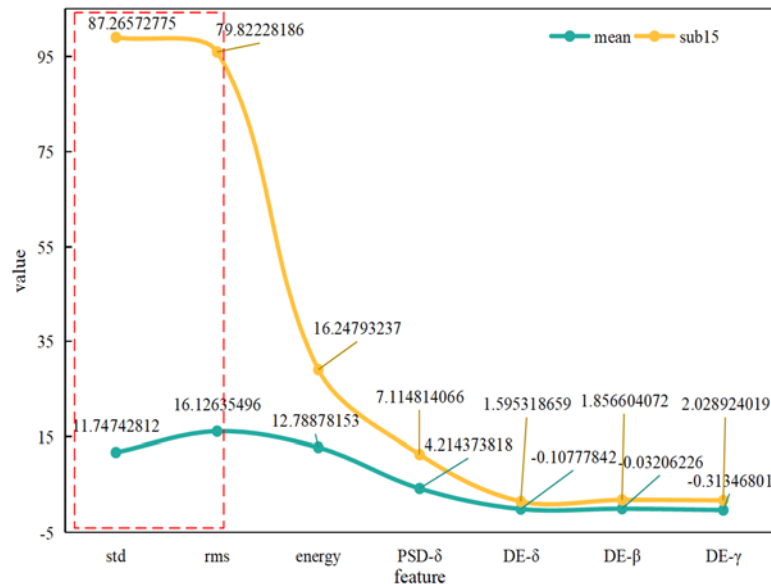

Figure 5. DEAP dataset misjudgment analysis comparison table.

From Figure 5, it could be seen that the SD and RMS of subject 15 showed significant differences from the mean of the same features of the training set. When SDmean = 11.75 and SD15 = 87.27, ΔSD ≈ 75.52. When RMSmean = 16.13 and RMS15 = 79.8, ΔRMS ≈ 63.69. Other feature values of subject 15 were also higher than the mean of the same features of training set. This indicated that subject 15 exhibited significant individual differences in the dataset, and was the reason why it was misjudged. The subject of individual differences should be included in the training set for training SSA-RF to obtain the better generalization ability.

3.4. Cross-subject emotion recognition on SEED dataset

We conducted 15 randomized grouping experiments, 12 subjects were randomly selected as the training set, the other 3 people were selected as the test set for 30 experiments, and 20 iterations were carried out in each experiment.

A total of 18 features were extracted: ZCR, SD, Mean, RMS, Eng, Skewness, Hjorth, PSD and DE of five frequency bands (δ, θ, α, β, and γ), and SE in the time-frequency domain. We performed many experiments of different feature combinations, and selected the top 8 combinations with high accuracy. These are shown as follows:

Combination 1: All time domain features;
Combination 2: All;
Combination 3: RMS, Eng, PSD-δ, DE-δ, DE-β, DE-γ;

Combination 4: DE-β, DE-α, PSD-β, PSD-α;
Combination 5: Eng, Skewness, Hjorth, PSD;
Combination 6: T-all and F-all;
Combination 7: DE-θ, DE-δ, PSD-θ, PSD-δ;
Combination 8: T-all and PSD;

The parameters value of RF were generally based on empirical evidence, and the empirical values of DTN and MLN were 30 and 1, respectively, but they were not suitable for each type of data. We applied SSA to automatically search for optimal parameters (DTN and MLN) of RF of the 8 combinations, and the optimal values of DTN and MLN were shown in Table 8.

Table 8. Optimization results of SSA-RF model parameters.

Feature combination Parameter	1	2	3	4	5	6	7	8
DTN	33	35	32	50	34	28	34	43
MLN	10	9	4	4	1	4	16	13

Since multiple experiments were performed for each feature combination of the SEED dataset, the results of each experiment were recorded and statistically analyzed to draw a violin plot as shown in Figure 6.
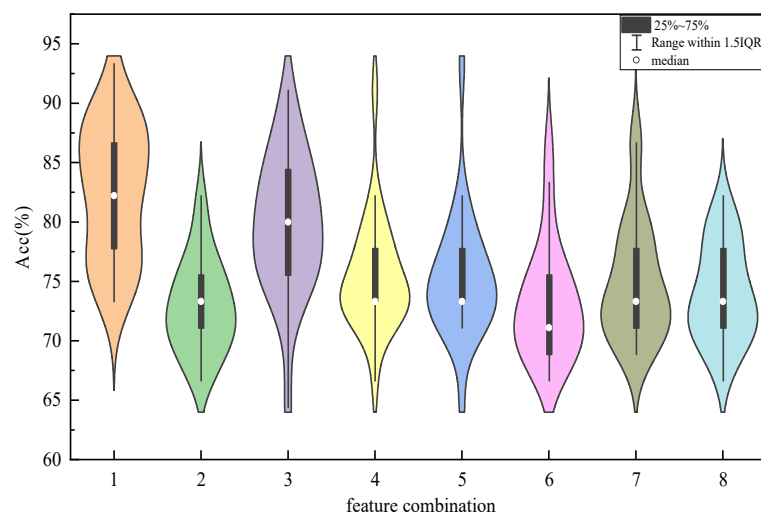

Figure 6. Results of different feature combinations of the SEED dataset.

It can be seen from the violin plot the accuracy corresponding to the feature combinations of the SEED dataset. The accuracy of the combination 1 was significantly higher than that of the other 7 combinations, and the accuracy of each combination was in the range of 65–93%, whose numerical span was larger than the DEAP dataset.

The experimental results of SSA-RF in the SEED dataset indicated that the accuracy of training set was close to 100%. The accuracy and improvement of the three classifications in the test set are shown in Figure 7 and Table 9.
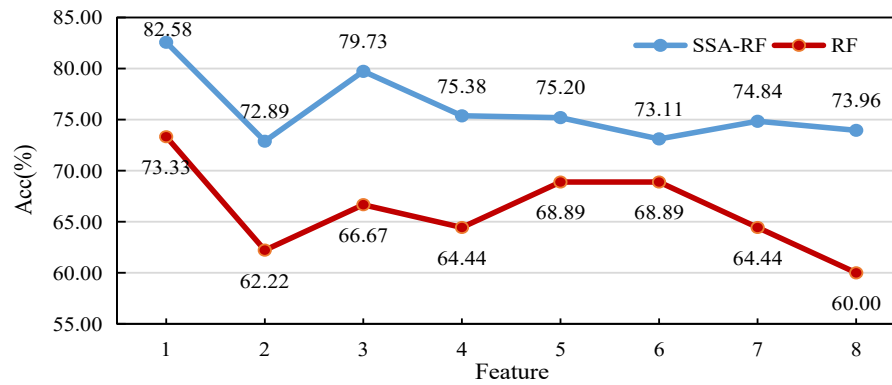
Figure 7. SSA-RF model test set classification results.

From Figure 7 and Table 9, it can be concluded that SSA-RF had a better optimization effect on the SEED dataset than the DEAP dataset. It can be seen that the accuracy of SSA-RF was higher than RF on each feature combination, with an average improvement of 9.85%. Among them, the accuracy of all time-domain feature combinations was 82.58%, with an improvement of 9.25%.

Table 9. The test set recognition accuracy comparison of SSA-RF and RF.

Feature combination Method (%)	1	2	3	4	5	6	7	8
SSA-RF	82.58	72.89	**79.73**	75.38	75.20	73.11	74.84	73.96
RF	73.33	62.22	**66.67**	64.44	68.89	68.89	64.44	60.00
Difference	9.25↑	10.67↑	**13.06↑**	10.94↑	6.31↑	4.22↑	10.40↑	13.96↑
Average				9.85↑				

For the misjudgment analysis of the discrimination results of combination 1 (All time domain features) in the SEED dataset, we extracted the feature data of combination 1 for subject 1 when misjudging (misjudging positive emotions as negative emotions), and compared it with the mean of those in the training set, as shown in Figure 8.

From Figure 8, it can be seen that the ZCR and SD of subject 1 shows significant differences compared with the mean of those in the training set.

When $ZCR_{mean} = 6930.1$, $ZCR_1 = 218,213.6$, which was almost a thirty-fold difference. Meanwhile, when $SD_{mean} = 2880.4$, $SD_1 = 206,475.2$, and the other feature values of subject 1 were also higher than the mean of those in the training set. Therefore it could be seen that subject1 exhibited significant individual differences in this dataset. For this reason, its accuracy was lower. Subsequent work needs to include subject 1 in the training set to train SSA-RF to obtain better generalization ability.
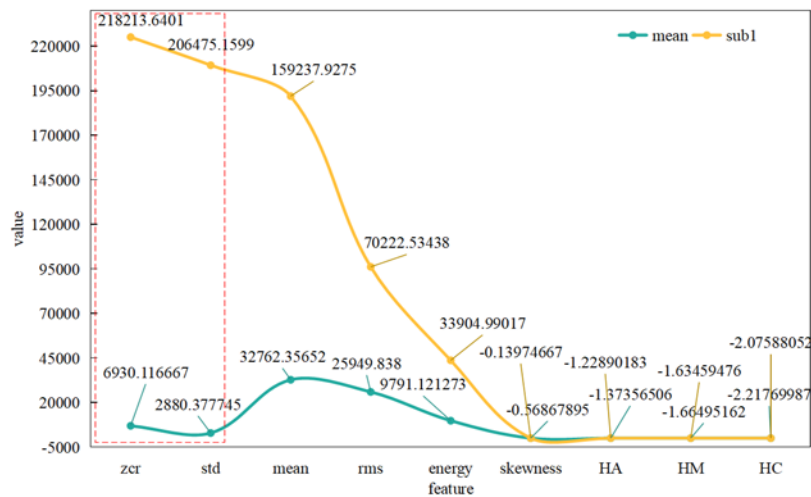
Figure 8. SEED dataset misjudgment analysis comparison table.

3.5. Comparison of similar methods

In the course of our research, we compared with particle swarm optimization algorithm (PSO), whale algorithm (WOA), and genetic algorithm (GA) with SSA algorithm, applied to the DEAP dataset. The experimental results are shown in Table 10.

Table 10. Comparison results of similar algorithms.

Feature Combination / Method (%)	1	2	3	4	5	6	7	8
PSO-RF	75.23	77.50	77.50	73.18	59.55	70.23	63.41	64.32
WOA-RF	71.36	54.09	73.18	69.77	65.68	59.55	69.55	61.36
GA-RF	62.73	72.50	74.55	74.09	63.41	74.55	62.05	68.40
SSA-RF	76.70	77.27	77.57	76.80	77.40	76.52	76.40	75.83
RF	73.05	76.88	74.55	75.03	76.59	76.19	73.86	75.39

To sum up, SSA has the best effect compared to similar algorithms.

4. Discussion

We compared the findings of this paper with previous research, the average accuracy was based on 100 epochs, and the results of the comparison are shown in Table 11.

We compared SSA-RF and the relevant references from the past 7 years in Table 11. The average accuracy of our method was 76.81%, which was higher than the others the on the DEAP dataset, The average accuracy of our method was 75.96%, which was higher than the others on the SEED dataset. SSA-RF improved the accuracy of cross-subject emotion recognition.

Table 11. Comparison with previous methods.

Dataset	References (Year)	Model	Average Acc (%)
DEAP	Arnau-Gonzalez et al. [28] (2017)	SVM	73.41
	Li et al. [29] (2018)	SVM	59.06
	Pandey et al. [30] (2019)	Deep Neural Network	62.50
	Cimtay et al. [31] (2020)	CNN	72.81
	Mert Ahmet et al. [32] (2021)	ANN	70.02
	Xu et al. [33] (2022)	GRU-Conv	70.07
	She et al. [34] (2023)	DDSA-mRMR-SRM	64.40
	Ours	**SSA-RF**	**76.81**
SEED	Lan et al. [35] (2018)	MIDA	72.47
	Gupta et al. [36] (2019)	Random forest classification model	72.07
	Luo et al. [37] (2020)	sWGAN + SVM	67.7
	Topic et al. [38] (2021)	TOPO-FM and HOLO-FM	73.11
	Emsawas et al. [39] (2022)	MultiT-S ConvNet	54.60
	Zhang et al. [40] (2023)	Semi-supervised emotion recognition model	73.26
	Ours	**SSA-RF**	**75.96**

5. Conclusions

At present, there has been no research on SSA optimizing RFs in the field of emotion recognition based on EEG. This research demonstrated that SSA-RF can obtain better accuracy in cross-subject emotion recognition. After extracting the composite domain features of EEG signals, we conducted a variety of feature combination experiments. Through this method, we found the optimal parameters of RF, and the accuracy was significantly improved. For the DEAP dataset, the average accuracy was 76.81%, with a maximum accuracy of 77.57%, which was 1.61% higher than RF. For the SEED dataset, the average accuracy was 75.96%, with a maximum accuracy of 82.58%, which was 9.85% higher than RF.

The SSA-RF algorithm proposed in our research is applicable to the classification training of personal emotion models, solving the inefficiency problems of high time cost and low adaptability of setting model parameters manually. SSA-RF can be applied in practice, and it has certain theoretical and practical significance for the development of emotion recognition.

Factors affecting the accuracy or efficiency of cross-subject emotion recognition also include baseline processing methods and automatic optimization of feature combinations. Therefore, multi-baseline processing and automatic optimization of feature selection have important research significance.

Author contributions

Conceptualization and methodology, X.Z.; software, S.W. and X.Z.; formal analysis, X.Z. and S.W.; investigation, X.Z. and Y.S.; data curation, K.X.; writing-original draft preparation, X.Z., S.W.;

writing-review and editing, X.Z. and K.X.; funding acquisition, X.Z. and R.Z. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Informed consent statement

The data used in this research are DEAP and SEED dataset which are the public datasets.

Data availability statement

http://www.eecs.qmul.ac.uk/mmv/datasets/deap/; https://bcmi.sjtu.edu.cn/~seed/index.html.

Conflict of interest

The authors declare no competing interests.

References

1. S. M. Alarcao, M. J. Fonseca, Emotions recognition using EEG signals: A survey, *IEEE Trans. Affective Comput.*, **10** (2017), 374–393. https://doi.org/10.1109/WiSPNET.2017.8299778

2. L. Piho, T. Tjahjadi, A mutual information based adaptive windowing of informative EEG for emotion recognition, *IEEE Trans. Affective Comput.*, **11** (2018), 722–735. https://doi.org/10.1109/TAFFC.2018.2840973

3. M. Kim, M. Kim, E. Oh, S. Kim, A review on the computational methods for emotional state estimation from the human EEG, *Comput. Math. Methods Med.*, **2013** (2013). https://doi.org/10.1155/2013/573734

4. R. H. Aljuhani, A. Alshutayri, S. Alahdal, Arabic speech emotion recognition from Saudi dialect corpus, *IEEE Access*, **9** (2021), 127081–127085. https://doi.org/10.1109/ACCESS.2021.3110992

5. Y. Liu, G. Fu, Emotion recognition by deeply learned multi-channel textual and EEG features, *Future Gener. Comput. Syst.*, **119** (2021), 1–6. https://doi.org/10.1016/j.future.2021.01.010

6. M. G. Salido Ortega, L. F. Rodríguez, J. O. Gutierrez-Garcia, Towards emotion recognition from contextual information using machine learning, *J. Ambient Intell. Hum. Comput.*, **11** (2020), 3187–3207. https://doi.org/10.1007/s12652-019-01485-x

7. R. Karbauskaitė, L. Sakalauskas, G. Dzemyda, Kriging predictor for facial emotion recognition using numerical proximities of human emotions, *Informatica*, **31** (2020), 249–275. https://doi.org/10.15388/20-INFOR419

8. B. Xie, M. Sidulova, C. H. Park, Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion, *Sensors*, **21** (2021), 4913. https://doi.org/10.3390/s21144913

9. Y. Li, B. Fu, F. Li, W. Zheng, A novel transferability attention neural network model for EEG emotion recognition, *Neurocomputing*, **447** (2021), 92–101. https://doi.org/10.1016/j.neucom.2021.02.048

10. H. Jiang, Z. Wang, R. Jiao, S. Jiang, Picture-induced EEG Signal classification based on CVC emotion recognition system, *Comput. Mater. Continua*, **65** (2020), 1453–1465. https://doi.org/10.32604/cmc.2020.011793

11. S. Zhang, C. Li, Research on feature fusion speech emotion recognition technology for smart teaching, *Mobile Inf. Syst.*, **2022** (2022). https://doi.org/10.1155/2022/7785929

12. Q. Liu, H. Liu, Criminal psychological emotion recognition based on deep learning and EEG signals, *Neural Comput. Appl.*, **33** (2021), 433–447. https://doi.org/10.1007/s00521-020-05024-0

13. X. Liu, S. Li, M. Wang, Hierarchical attention-based multimodal fusion network for video emotion recognition, *Comput. Intell. Neurosci.*, **2021** (2021). https://doi.org/10.1155/2021/5585041

14. J. Quan, Y. Miyake, T. Nozawa, Incorporating interpersonal synchronization features for automatic emotion recognition from visual and audio data during communication, *Sensors*, **21** (2021), 5317. https://doi.org/10.3390/s21165317

15. Y. Fang, H. Yang, X. Zhang, H. Liu, B. Tao, Multi-feature input deep forest for EEG-based emotion recognition, *Front. Neurorob.*, **14** (2021), 617531. https://doi.org/10.3389/fnbot.2020.617531

16. E. Anzai, D. Ren, L. Cazenille, N. Aubert-Kato, J. Tripette, Y. Ohta, Correction: Random forest algorithms to classify frailty and falling history in seniors using plantar pressure measurement insoles: a large-scale feasibility study, *BMC Geriatr.*, **22** (2022), 946. https://doi.org/10.1186/s12877-022-03425-5

17. T. Zhang, J. Su, Z. Xu, Y. Luo, J. Li, Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier, *Appl. Sci.*, **11** (2021), 543. https://doi.org/10.3390/app11020543

18. G. Beni, J. Wang, Swarm intelligence, in *Proceedings for the 7th Annual Meeting of the Robotics Society of Japan*, (1989), 425–428.

19. X. Ye, L. A. Dong, D. Ma, Loan evaluation in P2P lending based on random forest optimized by genetic algorithm with profit score, *Electron. Commerce Res. Appl.*, **32** (2018), 23–36. https://doi.org/10.1016/j.elerap.2018.10.004

20. J. K. Xue, B. Shen, A novel swarm intelligence optimization approach: sparrow search algorithm, *Syst. Sci. Control Eng.*, **8** (2020), 22–34. https://doi.org/10.1080/21642583.2019.1708830

21. W. Wang, F. Qi, D. P. Wipf, C. Cai, T. Yu, Y. Li, et al., Sparse Bayesian learning for end-to-end EEG decoding, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2023), 15632–15649. https://doi.org/10.1109/TPAMI.2023.3299568

22. X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, Y. Zhang, A survey on deep learning based brain computer interface: Recent advances and new frontiers, preprint, arXiv:1905.04149. https://doi.org/10.48550/arXiv.1905.04149

23. M. Z. Soroush, K. Maghooli, S. K. Setarehdan, A. M. Nasrabadi, Emotion recognition through EEG phase space dynamics and Dempster-Shafer theory, *Med. Hypotheses*, **127** (2019), 34–45. https://doi.org/10.1016/j.mehy.2019.03.025

24. H. Chao, L. Dong, Y. Liu, B. Lu, Emotion recognition from multiband EEG signals using CapsNet, *Sensors*, **19** (2019), 2212. https://doi.org/10.3390/s19092212

25. M. Shahbakhti, M. Beiramvand, I. Rejer, P. Augustyniak, A. Broniec-Wójcik, M.Wierzchon, et al., Simultaneous eye blink characterization and elimination from low-channel prefrontal EEG signals enhances driver drowsiness detection, *IEEE J. Biomed. Health. Inf.*, **26** (2021), 1001–1012. https://doi.org/10.1109/JBHI.2021.3096984

26. S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, et al., DEAP: A database for emotion analysis; using physiological signals, *IEEE Trans. Affective Comput.*, **3** (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

27. Y. Liu, Y. Ding, C. Li, J. Cheng, R. Song, F. Wan, et al., Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network, *Comput. Biol. Med.*, **123** (2020), 103927. https://doi.org/10.1016/j.compbiomed.2020.103927

28. P. Arnau-Gonzalez, M. Arevalillo-Herraez, N. Ramzan, Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals, *Neurocomputing*, **244** (2017), 81–89. https://doi.org/10.1016/j.neucom.2017.03.027

29. X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, B. Hu, Exploring EEG features in cross-subject emotion recognition, *Front. Neurosci.*, **12** (2018), 162. https://doi.org/10.3389/fnins.2018.00162

30. P. Pandey, K. R. Seeja, Subject independent emotion recognition from EEG using VMD and deep learning, *J. King Saud Univ. Comput. Inf. Sci.*, **34** (2022), 1730–1738. https://doi.org/10.1016/j.jksuci.2019.11.003

31. Y. Cimtay, E. Ekmekcioglu, Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition, *Sensors (Basel)*, **20** (2020), 2034. https://doi.org/10.3390/s20072034

32. A. Mert, H. H. Celik, Emotion recognition using time–frequency ridges of EEG signals based on multivariate synchrosqueezing transform, *Biomed. Eng./Biomed. Tech.*, **66** (2021), 345–352. https://doi.org/10.1515/bmt-2020-0295

33. G. Xu, W. Guo, Y. Wang, Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-Conv architecture, *Med. Biol. Eng. Comput.*, **61** (2023), 61–73. https://doi.org/10.1007/s11517-022-02686-x

34. Q. She, X. Shi, F. Fang, Y. Ma, Y. Zhang, Cross-subject EEG emotion recognition using multi-source domain manifold feature selection, *Comput. Biol. Med.*, **159** (2023), 106860. https://doi.org/10.1016/j.compbiomed.2023.106860

35. Z. Lan, O. Sourina, L. Wang, R. Scherer, G. R. Müller-Putz, Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets, *IEEE Trans. Cognit. Dev. Syst.*, **11** (2018), 85–94. https://doi.org/10.1109/TCDS.2018.2826840

36. V. Gupta, M. D. Chopda, R. B. Pachori, Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals, *IEEE Sensors J.*, **19** (2018), 2266–2274. https://doi.org/10.1109/JSEN.2018.2883497

37. Y. Luo, L. Z. Zhu, Z. Y. Wan, B. L. Lu, Data augmentation for enhancing EEG-based emotion recognition with deep generative models, *J. Neural Eng.*, **17** (2020), 056021. https://doi.org/10.1088/1741-2552/abb580

38. A. Topic, M. Russo, Emotion recognition based on EEG feature maps through deep learning network, *Eng. Sci. Technol.*, **24** (2021), 1442–1454. https://doi.org/10.1016/j.jestch.2021.03.012

39. T. Emsawas, T. Morita, T. Kimura, K. Fukui, M. Numao, Multi-kernel temporal and spatial convolution for EEG-based emotion classification, *Sensors*, **22** (2022), 8250. https://doi.org/10.3390/s22218250

40. Y. Zhang, Y. Peng, J. Li, W. Kong, SIFIAE: An adaptive emotion recognition model with EEG feature-label inconsistency consideration, *J. Neurosci. Methods*, **395** (2023), 109909. https://doi.org/10.1016/j.jneumeth.2023.109909